OPEN ACCESS Freely available online

PLOS ONE

Long Intergenic Non-Coding RNAs (LincRNAs) Identified by RNA-Seq in Breast Cancer

Xianfeng Ding[1]*, Limin Zhu[1], Ting Ji[1], Xiping Zhang[2], Fengmei Wang[3], Shaoju Gan[1], Ming Zhao[1], Hongjian Yang[2]*

1 Institute of Bioengineering, College of Life Science, Zhejiang Sci-Tech University, Hangzhou, Zhejiang, P.R. China, **2** Zhejiang Cancer Research Institute, Department of Breast Tumor Surgery, Zhejiang Cancer Hospital, Banshan Bridge, Hangzhou, Zhejiang, P.R. China, **3** Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, P.R. China

Abstract

In an attempt to find the correlation of aberrant expression of long intergenic noncoding RNAs (lincRNAs) with cancer, twenty-five samples of breast cancer tissue and respective adjacent normal tissue were studied for the expression of lincRNAs by RNA-seq. Among the 538 lincRNAs studied, 124 lincRNAs were exclusively expressed in cancer adjacent tissues and 62 lincRNAs were exclusively expressed in the cancer tissues. Furthermore, the expression of 134 lincRNAs was higher while 272 lower in breast cancer tissue compared with adjacent tissue. The expression of four selected lincRNAs (BC2, BC4, BC5, and BC8) was validated by semi-quantitative and real-time PCR. It was revealed that expression of lincRNA-BC5 was positively correlated with patients' age, pathological stage, and progesterone receptor concentration, while lincRNA-BC8 was negatively correlated with progesterone receptor expression. Higher expression of lincRNA-BC4 was seen in advanced breast cancer grade. LincRNA-BC2 showed no specific changes in the pathological features studied. Interactions between selected lincRNAs and breast cancer associated proteins were highly suggested by *RPIseq* based on the specific secondary structure. The results demonstrated that this group of lincRNAs was aberrantly expressed in breast cancer. They might play important roles in the function of oncogenes or tumor suppressors affecting the development and progression of breast cancer.

Citation: Ding X, Zhu L, Ji T, Zhang X, Wang F, et al. (2014) Long Intergenic Non-Coding RNAs (LincRNAs) Identified by RNA-Seq in Breast Cancer. PLoS ONE 9(8): e103270. doi:10.1371/journal.pone.0103270

Editor: Jin Q. Cheng, H. Lee Moffitt Cancer Center & Research Institute, United States of America

Received January 23, 2014; **Accepted** June 29, 2014; **Published** August 1, 2014

Funding: This work was supported by Natural Science Foundation of Zhejiang Province, China (No. LY12H16030); Foundation of Science and Technology Department of Zhejiang Province; China (No. 2011C23043) and also by Foundation of Health Department of Zhejiang Province, China (No. 2011BCA015). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: xfding@zstu.edu.cn (XD); yhjzlyy@163.com (HY)

Introduction

The human genome contains only 20,000 protein-coding genes, representing <2% of the total genome. Therefore substantial fractions of the human genome can be transcribed, yielding many short or long noncoding RNAs (ncRNA) with limited protein-coding capacity [1].

Long intergenic ncRNAs (lincRNAs) range in size from several hundred to tens of thousands of bases (≥200). They belong to a newly discovered class of ncRNAs. Although more than 3,000 human lincRNAs have been identified, less than 1% of them have been characterized [2]. Although still largely unexplored, ncRNA, particularly lincRNAs, have emerged as a new regulatory molecule exemplified by their frequent cell-type specific expression and subcellular compartment localization. They may play important roles in numerous systems, and interact with cancer related genes. Indeed, several well-described examples, such as HOTAIR [3], Xist [4], lincRNA-p21 [5], and MALAT-1 [6], indicate that lincRNAs may be essential factors in occurrences and developments of cancer [7,8].

Breast cancer is now considered a heterogeneous group of diseases with distinct clinical, pathological and molecular features. The latest report on cancer epidemiology indicated that breast cancer was the most common cancer in woman in 2013 [9]. Breast cancer is expected to account for 29% of all new cancer cases among women, and it ranked second in death rates. Recently, noncoding RNA, such as microRNAs [10] and lincRNAs [11], have become a hot topic in the development and progress of breast cancer. However, studies on lincRNAs in breast cancer are at a preliminary stage. Gupta reported that increased expression of HOTAIR was correlated with poor prognosis and tumor metastasis in breast cancer [3]. Recent studies showed that MALAT-1 was up-regulated in many human solid carcinomas, including lung, breast, colon, prostate and HCC [6]. However, the clinical value of lincRNA as an emerging group of ribonucleotides is still largely unknown in breast cancer. Although lincRNAs may have an impact on various human diseases [11], the detailed role and molecular mechanisms are still largely unknown.

Recent advance in RNA sequencing (RNA-seq) and computational methods allowed researchers to comprehensively annotate and characterize lincRNA transcripts [12,13]. Especially, paired-end RNA-seq, where 36–100 bp were sequenced from both ends of 200 to 500 bp long DNA molecules, was suitable for the detection of low-copy and novel lincRNAs [14].

To better understand the roles of lincRNAs in breast cancer development and progression, comprehensive analysis of the expression abundance of lincRNAs is required. In this study, we

described a comprehensive analysis of lincRNAs in twenty five pairs of snap-frozen breast cancer tissues and matched cancer adjacent tissues by RNA-Seq. We found that a group of lincRNAs was aberrantly expressed in twenty five breast cancer tissues compared with matched adjacent tissues. Five samples randomly chosen from these patients were analyzed for the expression of lincRNAs, by deep sequencing technology. Twenty samples from breast cancer patients were evaluated by real-time qPCR. The correlation between the expression levels of the selected lincRNAs with pathological parameters was analyzed. In order to further study the functional mechanism, secondary structures of the lincRNAs were projected by software *RNAfold* Web Server. Meanwhile, the possibility of interaction between lincRNAs and oncogenes was evaluated by *RPIseq*.

Results

RNA-Seq and Reads mapping

PolyA-minus RNAs were fractionated from total RNA samples isolated from pooled breast cancer tissues and matched adjacent tissues. Then, RNA-seq libraries were generated by RNA-fragmentation, random hexamer-primed cDNA synthesis, linker ligation and PCR amplification. The purified DNA libraries were sequenced by using Illumina Hi-seq 2000 platform.

A total of 27 million reads were obtained. 8.9 million-reads were from cancer tissues, and 18.1 million from the adjacent tissues (Figure 1). The obtained reads were first mapped to the human reference genome using the TopHat program (v1.0.3). Reads mapped to genome were then filtered against the RepeatMask and Ensembl gene sets, which resulted in the identification of novel intergenic transcription regions. These reads were then compared against rRNA and other repeated sequences. Finally there were 5.41 million and 13.4 million reads in breast cancer tissues and adjacent tissues, respectively after depletion of rRNA database matching with BOWTIE (0.12.7). 17.9% and 13.07% were mapped to the NONCODE 2.0 using TopHat (1.2.0) in cancer tissues and adjacent tissues respectively (Figure 1).

Bioinformatics analysis

For each transcription region, a FPKM (fragment per kilobase of transcript per million mapped reads) value was calculated to quantify its expression abundance and variations, using cufflinks v1.0.3.

To accurately define the boundaries of these intergenic transcript regions, two parameters were considered: the minimum distance from the two neighboring genes and the minimum number of mapped reads within a genomic region. The minimum distance was set 1500 nucleotides after the terminal of the upstream gene and also 1500 nucleotides before the origin of downstream gene. 10 mapped reads were considered as the standard minimum number. Based on these two principles, 538 long intergenic noncoding RNAs (lincRNAs) were identified with consistency (Table S9). Among them, there were 124 lincRNAs (Table S1) exclusively expressed in the tissues adjacent to cancer (FPKM>10) and 62 lincRNAs (Table S2) exclusively expressed in the cancer tissues (FPKM>10). Furthermore, 352 overlapped lincRNAs in both cancer and adjacent tissues were aberrantly expressed through calculating the fold-change of RPKM (the absolute value by log2 (reads in cancer tissues/reads in adjacent tissues to cancer)≥1). 134 lincRNAs were expressed higher (Table S3) and 272 lincRNAs (Table S4) were expressed lower in breast cancer tissues compared with adjacent breast cancer tissues. Considering their diverse functions in carcinogenesis, several most different lincRNAs from both up-regulated subgroup and down-regulated subgroup were selected to be further studied. This selection method was based on four principles. Firstly, the selected lincRNA were stability expressed in all the samples we ever used in our study. Secondly, the results of semi-quantitation PCR preliminary indicated the differences between breast cancer tissues and adjacent tissues. Thirdly, forming stable scaffold structure and having entropy for the lincRNA are two important factors. Fourthly, the clear locations of lincRNAs and gene on both sides are considered to screening lincRNAs for further study. The selected lincRNAs were named by the abbreviation of breast cancer (BC), such as lincRNA-BC2 (log2 Ratio = 2.46) and lincRNA-BC5 (log2 Ratio = 2.22) were the most significantly up-regulated lincRNAs, and lincRNA-BC4 (log2 Ratio = −2.64) and LincRNA-BC8 (log2 Ratio = −2.017) were the most significantly down-regulated ones.

Individual validation of lincRNAs by RT-PCR and qPCR

The solexa results from two pooled samples of five breast cancer tissues and matched adjacent cancer tissues were further validated individually by semi quantitative and real-time PCR. There were twenty (all the numbers are confusing and did not make sense in the current description) breast cancer tissues and matched adjacent tissues used in these experiments. The expression of these selected lincRNAs was validated by using RT-PCR and qPCR (Figure 2). Selected lincRNAs (Table S5) were validated by RT-PCR and the information of primers was provided in Table S6.

LincRNAs expression and clinical pathologic feature of breast tissues

The expression of selected lincRNAs was analyzed in 20 breast cancer tissues and matched adjacent tissues. We analyzed lincRNAs whose expression was significantly different between cancer tissues and adjacent tissues. LincRNA-BC2 (6.315±0.672, P = 0.00) and lincRNA-BC5 (2.72±0.46, P = 0.001) were consistently up-regulated more than 2-fold (mean±SD) in cancer samples. Whereas, lincRNA-BC4 (0.358±0.062, P = 0.00) and lincRNA-BC8 (0.436±0.0732, P = 0.00) were down-regulated ($p < 0.01$) (Figure 3A.). The consistent differential expression suggested that they could potentially act as tumor oncogenes or suppressor genes, respectively. We also analyzed the correlation between lincRNA expression and the clinical feature of the cancer including the ages of original diagnosis (median 54 years, range 32–73 years), cancer staging (grade II or III), lymph node metastasis (positive or negative), ER/PR (estrogen and progesterone receptor level), the level of HER2 (positive or negative) and p53 protein (positive or negative).

Expression of LincRNA-BC4 (P = 0.03) was significantly lower in grade III breast cancers compared to that in grade II cancers. On the contrary, the expression of lincRNA-BC5 (P = 0.007) was significantly higher in grade III. Meanwhile, there were no significant differences of lincRNA-BC2 (P = 0.301) and lincRNA-BC8 (P = 0.441) on the different grades in pathology (Figure 3B).

The expression of the lincRNAs (BC2, BC4, BC8) did not exhibit significant differences between ER positive (ER+) patients and ER negative(ER-)patients (P = 0.502, 0.438 and 0.344 respectively). Only lincRNA-BC5 (P = 0.00) expression was significantly lower in estrogen receptor negative patients compared to positive patients (Figure 3C). LincRNA-BC5 (P = 0.031) expression was also significantly lower on the progesterone receptor (PR) negative patients. The expression of lincRNAs BC2, BC4, and BC8 exhibited no difference between PR negative and PR positive (P = 0.782, 0.568 and 0.642 respectively) (Figure 3D).

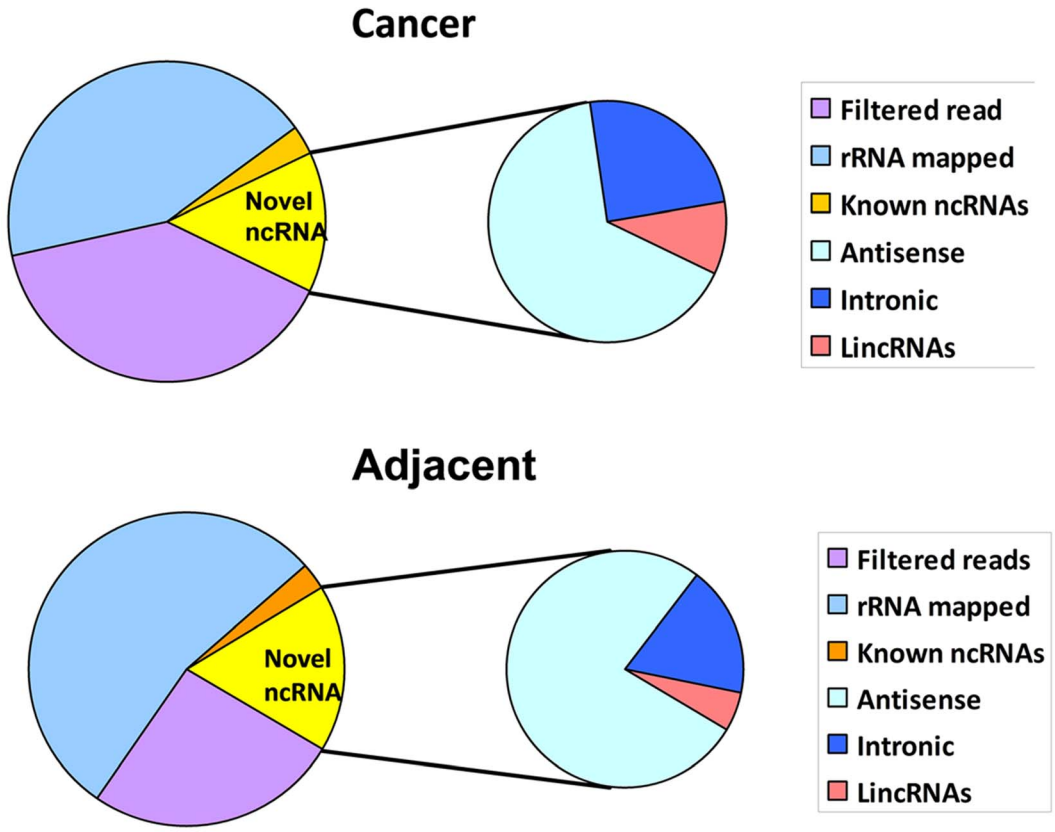

Figure 1. Global overview of polyA-minus RNA sequencing in breast cancer tissues and matched adjacent cancer tissues. The pie charts on the left display ploy-A minus transcript distribution in breast cancer tissues (upper) and adjacent tissues to cancer (lower). The pie charts on the right display novel ncRNA categorized as sense (intronic RNA and lincRNA) and antisense transcript.
doi:10.1371/journal.pone.0103270.g001

The expression of lincRNAs were not significantly different between HER2-negative and HER2-positive breast cancer tissues (for lincRNA-BC2, $P = 0.542$; for lincRNA-BC4, $P = 0.866$; for lincRNA-BC5, $P = 0.176$; lincRNA-BC8, $P = 0.166$) (Figure 4A). However, lincRNA-BC8 ($P = 0.004$) showed significantly higher expression in the p53-positive cancer tissues compared to the p53-negative ones. LincRNA-BC2 ($P = 0.526$), lincRNA-BC4 ($P = 0.867$), and lincRNA-BC5 ($P = 0.393$) were not significantly different between p53 positive and p53 negative breast cancer tissues (Figure 4B).

The expression of lincRNA-BC2 ($P = 0.017$) was significantly higher in patients with lymph node metastasis compared to patients without metastasis. The expression of other lincRNAs, BC4, BC5 and BC8 were not significantly different between the metastatic and non-metastatic patients. In terms of age, lincRNA-BC5 ($P = 0.021$) was little expressed in older group (>54 years) compared to younger group (≤54 years). LincRNA-BC2($P = 0.502$), lincRNA-BC4($P = 0.438$), and lincRNA-BC8($P = 0.344$) did not show statistically significant difference between the two groups. (Figure 4D). The information of qPCR primers was provided in Table S7, and the average levels of lincRNAs expression and correlation with clinic pathology was presented in Table S8.

The qPCR result showed that there was a significant correlation between lincRNAs and carcinogenesis. LincRNA-BC5 was specially related with high degree of differentiation and elder age. However, LincRNA-BC2 was a factor unrelated to hormone level, but more highly expressed in breast cancer. It was suggested that lincRNA-BC2 might play a role in triple negative breast cancer.

Validation of chromosomal location

In order to prove the accuracy of the selected lincRNAs, the "intergenic" positions need to be validated precisely. The transcription initiation sites and termination sites of bilateral genes and lincRNAs were exactly mapped on chromosomes. We validated the position of the four lincRNAs on chromosomes through BLASTN alignments from NCBI Genebank. Each "intergenic" region and adjacent gene was confirmed. For lincRNA-BC2 (chr5q33: 149,876,146–149,876,368), the upstream gene was RPS14 (40S ribosomal protein S14), and the downstream gene was NDST1 (N-deacetylase/N-sulfotransferase (heparin glucosaminyl). LincRNA-BC4 (chr15q21: 49,983,192–49,938,395) was between DTWD1 (DTW domain containing 1) and RLIMP3 (ringer finger protein, LIM domain interaction pseudogene 3). LincRNA-BC5 (chrXq24: 114,962,257–114,962) was between Loc728825 small ubiquitin-like modifier and AKRIBP8 (aldo-keto reductase family 1 member B1 pseudogene). For lincRNA-BC8 (chr13q34: 110,076,492–110,076,722), the bilateral genes were MY016-AS1 (MY016 antisense RNA1) and IRS2 (Insulin receptor substrate 2). The loci information was consistent with the result from bio-informatics analysis. The transcriptional initiation sites and termination sites of neighbor genes on chromosome were shown in Figure 5.

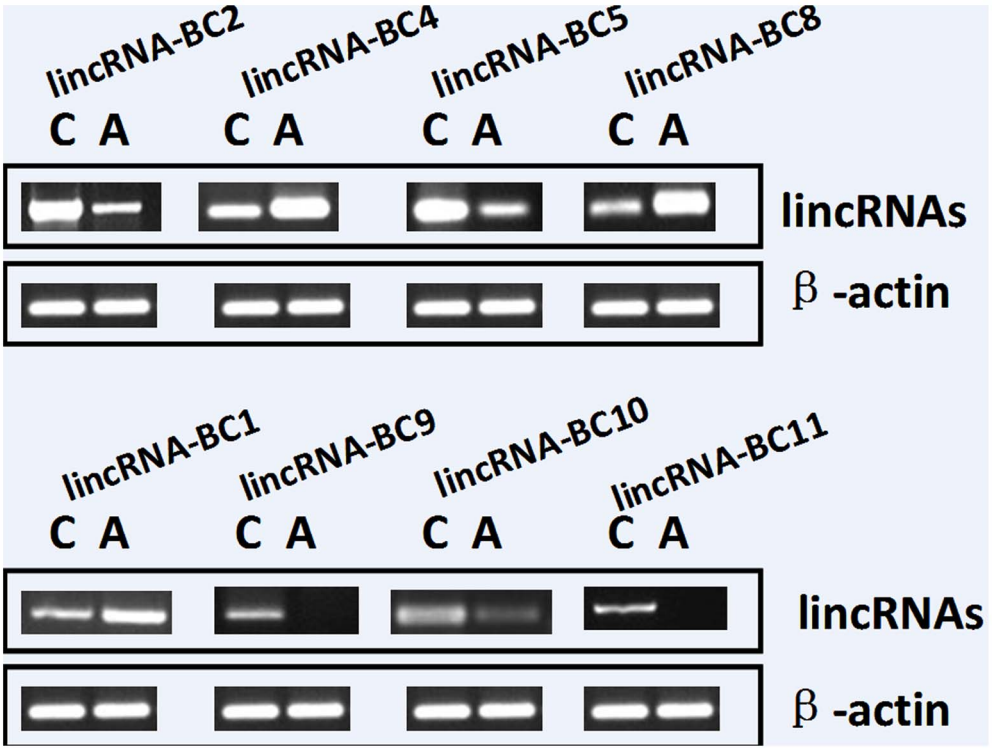

Figure 2 The expression of selected lincRNAs in breast cancer tissues and matched adjacent tissues to cancer using semi-quantitative RT-PCR.
doi:10.1371/journal.pone.0103270.g002

Prediction of LincRNAs' secondary structure

RNA sequences are single-strand biopolymers which can fold themselves. The potential interactions in organism are determined by RNA secondary structure [15]. The prediction of RNA structure can be the first important step for the functional characteristics of novel lincRNAs [16]. The structure of the selected lincRNA were predicted with computer software *RNAfold* Web Server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

The results showed that they all had low free energy from algorithms. Meanwhile, there were more than three stem-loop structures by self-fold (Figure 6). This specific structure means that there probably were proteins or chromosomes binding sites. It was suggested that the selected lincRNAs may be involved in the complex chromatin-modifying complexes or in the regulation of gene transcription.

RPISeq prediction

RNA-protein interactions play important roles in a wide variety of cellular processes, ranging from transcriptional and post-transcriptional regulation of gene expression to host defenses against pathogens [17]. RNA-Protein interaction prediction was performed by *RPIseq*. *RPISeq* (http://pridb.gdcb.iastate.edu/RPISeq/), a family of purely sequence-based classifiers, can be used to predict whether a specific RNA-protein is likely to interact. Two variants of *RPISeq* were presented:*RPISeq-SVM* (Support Vector Machine (SVM) classifier) and *RPISeq-RF* (Random Forest classifier). Predictions with probabilities >0.5 were considered positive.

The results showed that all four lincRNAs had great possibility of interaction with BRCA1 and BRCA2 (Figure 7). For example, LincRNA-BC2 was predicted to interact with BRCA1 by RF (P = 0.55) and SVM (P = 0.779). It was also predicted to interact with BRCA2 by RF (P = 0.90) and SVM (P = 0.772). All of the four lincRNAs were predicted to interact with ER or HER2 by scores of RF and SVM bigger than 0.5. It was therefore suggested that these four lincRNAs may involve in the occurrence of breast cancer since ER and HER2 had been used as clinical biomarkers for diagnosis. The other proteins targeted for potential interaction with the lincRNAs were listed in the Figure 6. However, there were different results of prediction for interaction of lincRNA-BC8 with proteins by RF classifier (0.4–0.65) and SVM classifier (0.886–0.989).

Discussion

The development of high throughput deep sequencing technology provided the possibility of a nearly complete view of lincRNAs profiles [18]. Deep sequencing technology had the potential to identify novel tissue-specific lincRNAs. This new technology had the advantages of providing not only sequence of low abundance species, but also quantitative data since the frequency of sequencing reads reflects the abundance of lincRNAs in the population.

Recent studies demonstrated that lincRNAs are exquisitely regulated during the development of cancer. They responded to diverse signaling cues and were aberrantly expressed in diverse cancers tissues.

In this study, deep sequencing technology was used to detect the expression profiles of lincRNA in five pairs of snap-frozen breast cancer tissues and matched adjacent cancer tissues. Expression of many lincRNAs was significantly altered in cancer tissues compared with matched adjacent tissues, suggesting that these aberrantly expressed lincRNAs might play roles in carcinogenesis. Real-time qPCR was performed to evaluate the expression pattern

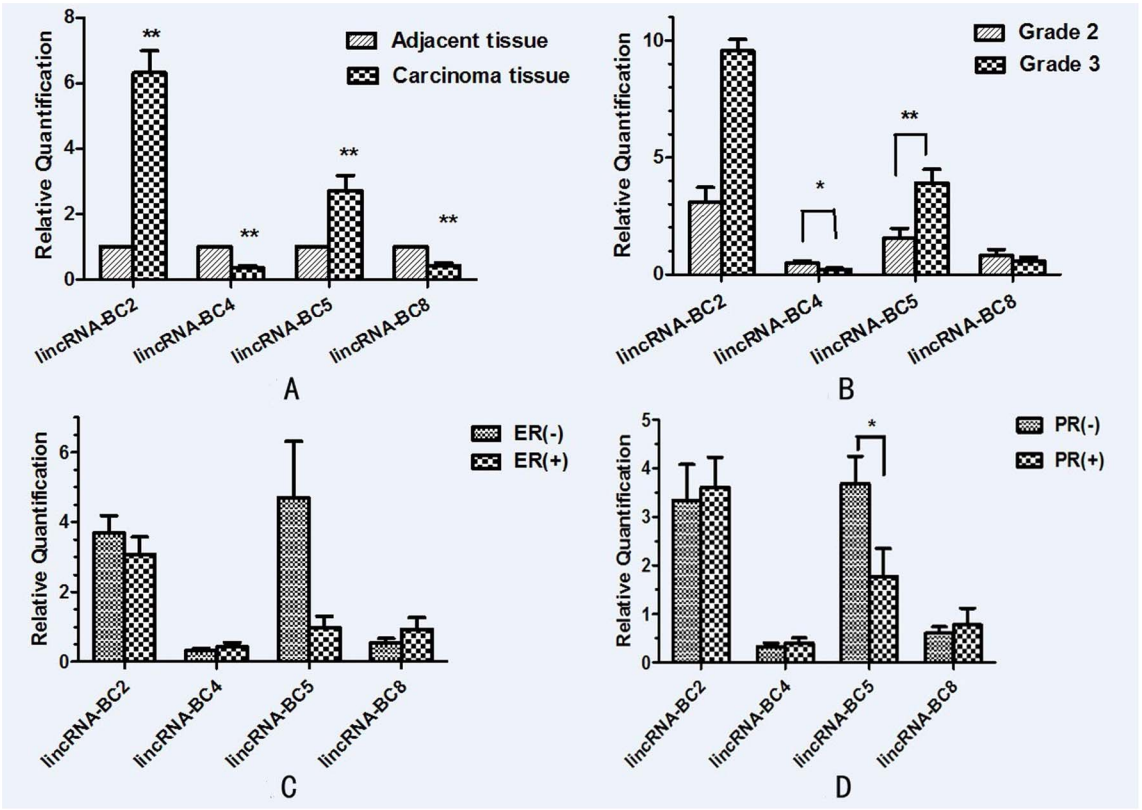

Figure 3. Comparison of lincRNAs (BC2, BC4, BC5, BC8) aberrant expression in breast cancer tissues with matched adjacent tissues to cancer with clinic pathological. (A) Comparison of lincRNA aberrant expression in breast cancer tissues with matched adjacent tissues to cancer. (B) Comparison of lincRNA aberrant expression in breast cancer tissues on tumor grades between grade II and grade III. (C) Comparison of lincRNA aberrant expression in breast cancer tissues on the level of ER (estrogen receptor). (D) Comparison of lincRNA aberrant expression in breast cancer tissues on the level of PR (progesterone receptor).
doi:10.1371/journal.pone.0103270.g003

of lincRNA-BC2, lincRNA-BC4, lincRNA-BC5, and lincRNA-BC8 in twenty carcinoma patients. LincRNA-BC2 and lincRNA-BC5 expression were upregulated while; lincRNA-BC4 and lincRNA-BC8 expression were down-regulated in breast cancer tissues compared to matched adjacent tissues. These findings were consistent with the results from deep sequencing analysis.

We used the matched adjacent tissues as control against the cancer tissue in our study. However, it was previously reported that preliminary changes on transcription took place in tissues adjacent to the carcinoma. Due to the ethical difficulties of obtaining normal breast tissues, β-actin housekeeping gene was used as internal control gene in the validation phase [19]. Expression of β-actin was relatively constant in both breast cancer tissues and adjacent tissues. Therefore β-actin could be used as a control normalizer in real-time quantitative PCR.

The function of the selected lincRNAs (lincRNA-BC2, BC4, BC5, BC6) remain largely unknown but the result demonstrated that they were aberrantly expressed in cancer tissues compared to adjacent cancer tissues by qPCR. Especially the expression level of the lincRNA-BC5 was positively correlated with patients' age, pathological stage, and progesterone level which were statistically significant.

The locations of the four studied lincRNAs were also extremely important. For the position of lincRNA-BC2, Dhillon et al. indicated that chr5q33 as a fragile site may be the unstable sites in the genome and can be used as suitable and reliable markers for genetic factor to breast cancer, epithelial ovarian cancer, and in non-small-cell lung cancer [20]. LincRNA-BC4 was mapped to chr15q21. Recent studies provided increasing evidence that chromosomal arm 15q may be an important target of genetic alterations in the progression of breast cancer [21]. The expression of chr15q21 was proved to be related with the risk of Griscelli syndrome [22], chronic lymphocytic leukemia [23], and prostate cancer. LincRNA-BC5 was located on chromosome Xq24. This region was reported to be associated with height [24]. In addition, the transcription level of this region might influence the risk level of obesity [25]. LincRNA-BC8 was mapped to Chr13q34 which was quite an interesting location. Lorenzo et al [26] reported that 13q34 amplification was a genomic aberration, and it was associated with basal-like breast cancer. Furthermore, this amplification had been previously reported in squamous cell carcinomas [27], adrenocortical carcinomas [28], childhood medulloblastoma [29], hepatocellular carcinomas [30] and breast cancer [26]. These lincRNAs selected from the solexa data of breast cancer tissues were mapped to disease-associated loci. Further studies are required to discover their important roles in carcinogenesis.

LincRNAs target chromatin modification complexes or RNA-binding protein to alter gene expression programs. One of the well-characterized lincRNAs is HOTAIR, which is transcribed between the HOXC clusters and represses genes in the HOXD cluster by binding and recruiting the chromatin-modifying complex PRC2. LincRNAs may carry out lots of functions by acting as modular scaffolds for protein-chromatin interactions

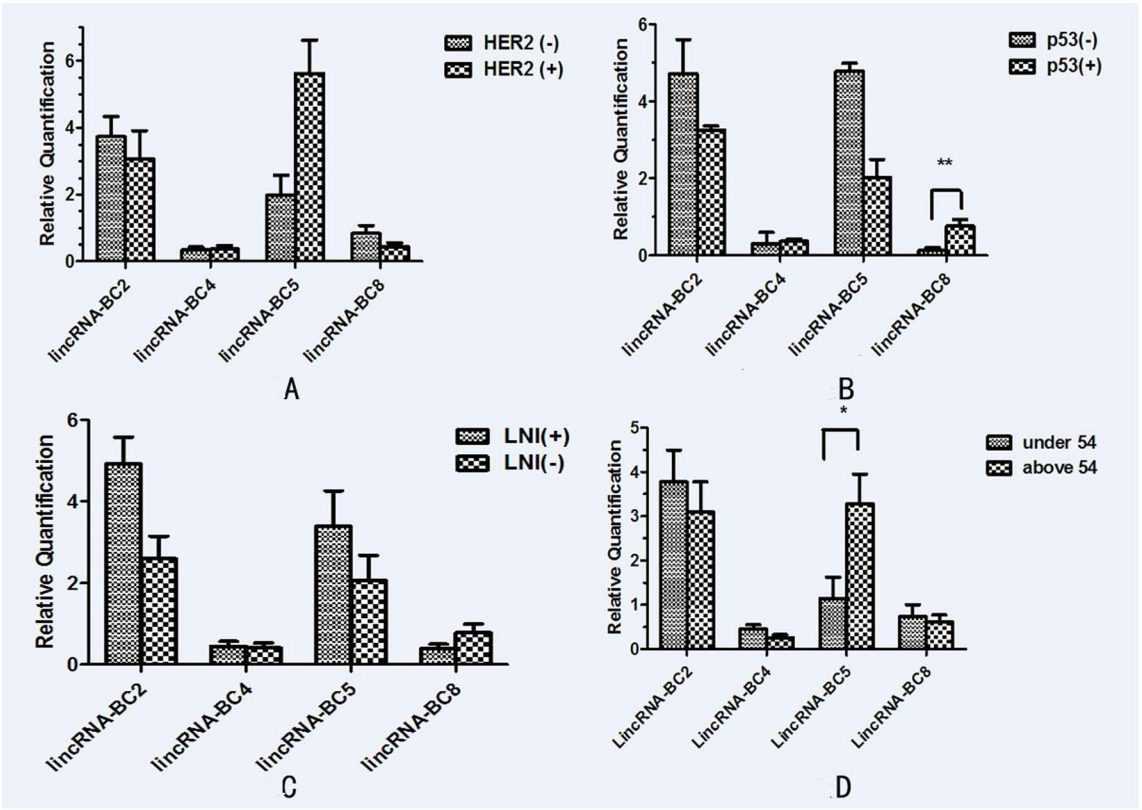

Figure 4. Comparison of lincRNA aberrant expression in breast cancer tissues with matched adjacent tissues to cancer in clinic pathological. (A) Comparison of lincRNA aberrant expression in breast cancer tissues with matched adjacent tissues to cancer on the level of HER-2. (B) Comparison of lincRNA aberrant expression in breast cancer tissues on the level of p53. (C) Comparison of lincRNA aberrant expression in breast cancer tissues on the level of LNI (lymph node metastasis). (D) Comparison of lincRNA aberrant expression in breast cancer tissues on the age of patients ($<$54 and $\geq$54).
doi:10.1371/journal.pone.0103270.g004

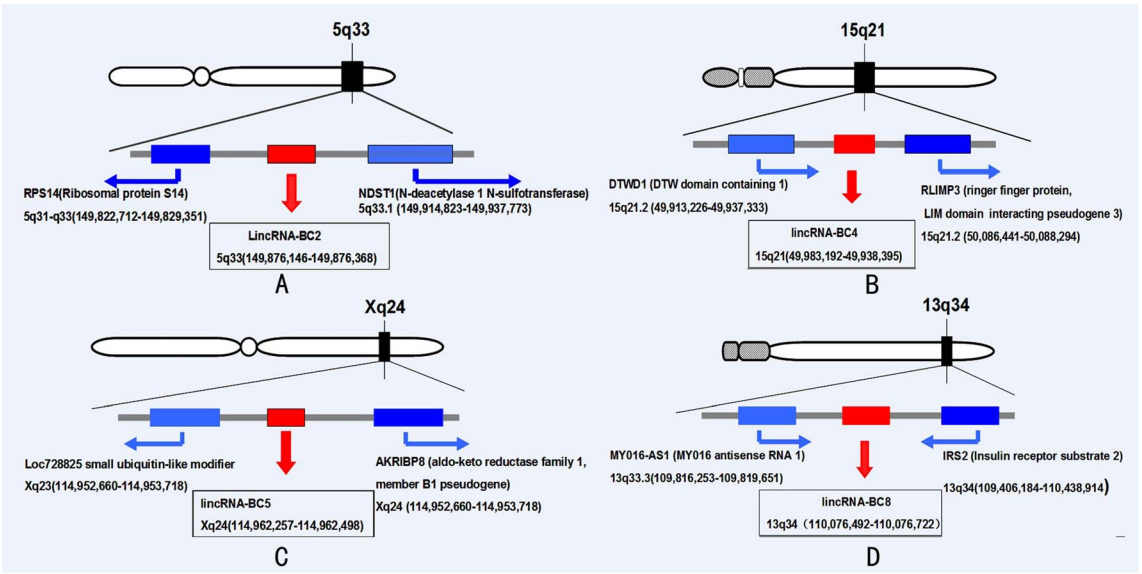

Figure 5. The position of lincRNAs and neighboring genes in chromosome. (A) The position of lincRNA-BC2 and neighboring genes in chromosome 5. (B) The position of lincRNA-BC4 and neighboring genes in chromosome 15. (C) The position of lincRNA-BC5 and neighboring genes in chromosome X. (D) The position of lincRNA-BC8 and neighboring genes in chromosome 13.
doi:10.1371/journal.pone.0103270.g005

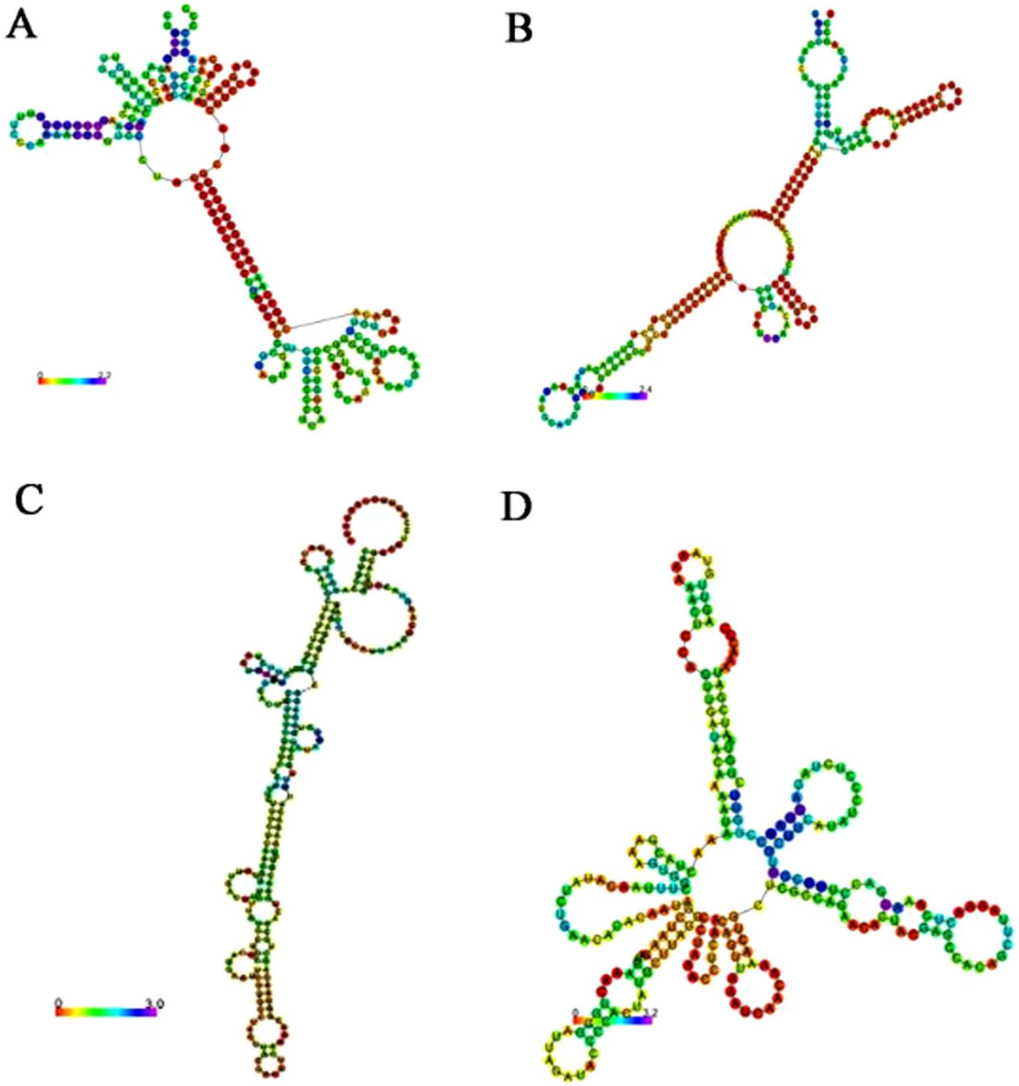

Figure 6. The second structure predicted by *RNAfold*. (A) The second structure of lincRNA-BC2 predicted by *RNAfold*. (B) The second structure of lincRNA-BC2 predicted by *RNAfold*. (C) The second structure of lincRNA-BC2 predicted by *RNAfold*. (D) The second structure of lincRNA-BC2 predicted by *RNAfold*.
doi:10.1371/journal.pone.0103270.g006

[31]. Tsai et al. suggested that lincRNAs may serve as scaffold by providing binding surfaces to assemble selected histone modification enzymes, thereby specifying the pattern of histone modification on target genes. LincRNAs play multiple functions with specific structure of several stem loops. In this study, the presence of multiple binding sites enables the RNA to specifically associate with DNA, RNA, and/or protein.

The precise mechanism of lincRNAs function remains poorly understood. However, one emerging theme is the interaction between lincRNAs and proteins. The functional importance of many lincRNA-protein interactions in transcriptional regulation has been demonstrated, such as XIST,HOTAIR and lincRNA-p21. *RPIseq* was used to predict the interaction between lincRNAs (BC2, BC4, BC5 and BC8) and breast cancer associated protein. *RPIseq* was a reliable method using only sequence-derived information [17]. Davide et al. had used this method to study the probabilities of interactions between HOTAIR and Suz12 [32].

We analyzed the interaction between lincRNAs and breast cancer associated proteins including BRCA-1 [33], BRCA-2 [34], PR, ER [35], p53, HER2 [36], K-ras [37], PTEN [38], TNF [39], and EGRF. All of the ten proteins were identified to play important roles in development, especially in the progress of tumor invasion and metastasis [40]. *RPIseq* is a sequence-based predictive method. The accuracy of the prediction ranged from 57–99% [17] in independent datasets of RNA-protein interactions. We found that most of scores were more than 0.5 analyzing interaction probabilities between the four lincRNAs and the proteins. It suggested that the four lincRNAs may participate in the carcinogenesis. However, computational prediction of lincRNA functions is still at its primary stage. More softwares were required to confirm the potential interactions of lincRNA. In addition, experimental methods such as RNA Knocking-Out, Western Blotting should be adopted in the future research.

Our data provides novel insight into breast cancer biology. A collection of lincRNAs was aberrantly expressed in breast cancer, suggesting that they might play roles as oncogenes or tumor

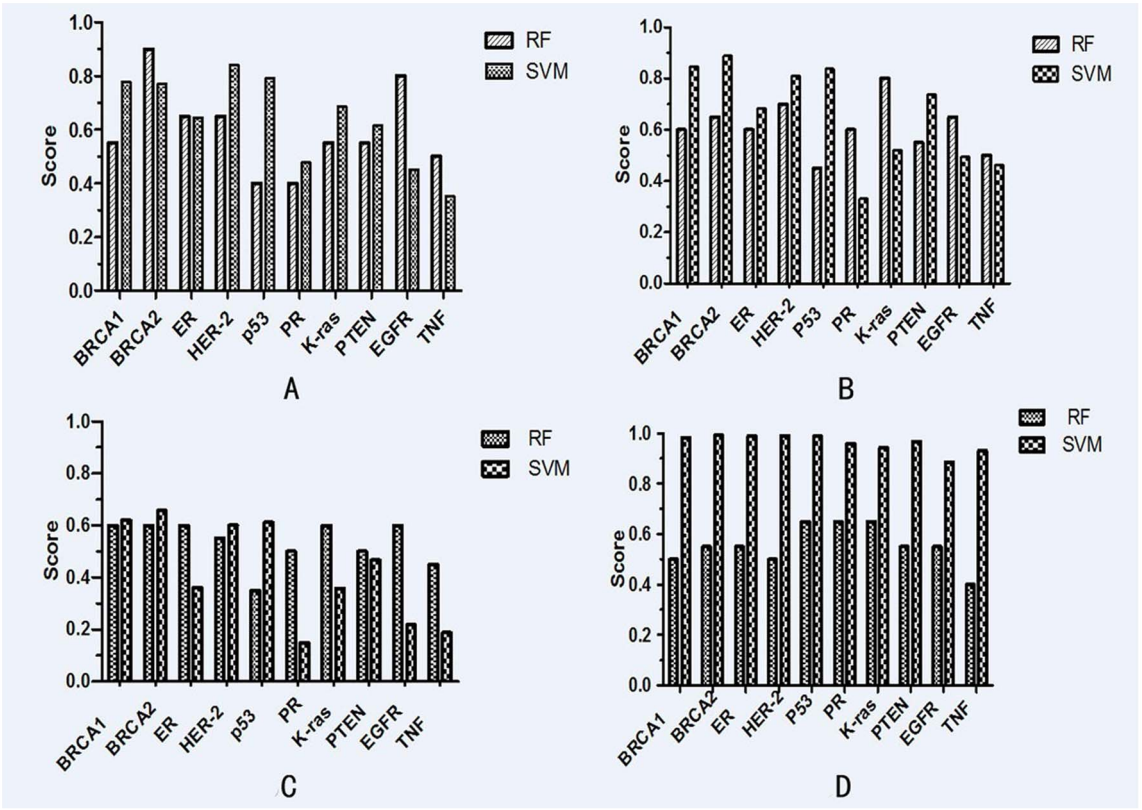

Figure 7. The scores of the interaction probability between lincRNAs and breast cancer associated protein predicted by *RPIseq*. (A) The scores of the interaction probability between lincRNA-BC2 and breast cancer associated protein predicted by *RPIseq*. (B) The scores of the interaction probability between lincRNA-BC4 and breast cancer associated protein predicted by *RPIseq*. (C) The scores of the interaction probability between lincRNA-BC5 and breast cancer associated protein predicted by *RPIseq*. (D) The scores of the interaction probability between lincRNA-BC8 and breast cancer associated protein predicted by *RPIseq*.
doi:10.1371/journal.pone.0103270.g007

suppressors in the development and progression of cancer. But it is necessary that large amount of breast cancer samples need to be collected to verify our results. Therefore, more work will be done to determine potential functions and regulatory mechanism of lincRNAs in breast cancer.

Materials and Methods

Ethics statement

Tissues from twenty-five women with breast cancer were collected from Zhejiang Cancer Hospital (time from February of 2008 to December of 2010). This study was approved by Zhejiang Provincial Experimental Animal Manage Committee under Contract 2013–2069 (ZEAC 2013–2069). All aspects of the study comply with the Declaration of Helsinki. All the samples were collected with informed consent of the patients. Ethics Committee specifically approved that not informed consent was required because data were going to be analyzed anonymously. Furthermore, there is no security and privacy violation to the patient's health in our study.

Preparation of patients' samples

None of the patients had received any radiotherapy and/or chemotherapy. All the patients were diagnosed as infiltrating carcinoma by pathology. Fresh tissues were harvested from patients, snap-frozen, and preserved at $-80^{\circ}C$ until further use. Clinical and pathological parameters of patients were recorded once diagnosed with breast cancer, including the value of ER, PR, HER-2. p53 and the age of patients. Patient characteristics were summarized in Table 1. Twenty-five breast cancer tissues and their corresponding adjacent tissues to cancer from patients were collected. Of these patients, five were used for initial deep sequencing analysis of lincRNAs and twenty were used for validation by qPCR.

PolyA-minus RNA preparation and next generation sequencing

Total RNA was isolated from breast cancer tissues and adjacent tissues using miRNeasy Kit according to the manufacturer's instructions. An additional DNase I digestion step was performed to ensure that the samples were not contaminated with genomic DNA. RNA purity was assessed using the Nanodrop-2000. Each RNA sample had an A260:A280 ratio above 1.8 and A260:A230 ratio above 2.2.

Five μg of breast cancer tissues were subjected to ribosomal RNA depletion according to the manufacturer's protocol of Ribominus kit. Then RNA was fragmented into ~200 base pairs (bp) and quantified with Nanodrop. The cDNA libraries were generated by RNA fragmentation, random hexamer-primed cDNA synthesis, linker ligation and PCR amplification.

cDNA was then used for Illumina sequencing library preparation. DNA fragment (200 ng) was then end-repaired to generate blunt ends with 5′ phosphatase and 3′ hydroxyls and adapters were ligated for paired end sequencing on Illumina HiSeq 2000.

Table 1. Clinic pathologic characteristics of patients with breast cancer.

Variable	Clinic pathologic parameter	Number of cases	Number of cases for qPCR
Case		25	20
Age	≤54	13	10
	>54	12	10
Therapy	no	25	20
	Chemotherapy	0	0
	Radiotherapy	0	0
Histological grade [a]	I	0	0
	II	10	10
	III	15	10
ER	negative	14	13
	positive	11	7
PR	negative	15	12
	positive	10	8
p53	negative	3	3
	positive	22	17
HER2	negative	15	13
	positive	10	7
Lymph node metastasis	no	11	13
	1	14	0
	≥1	0	7
	unknown	0	0

[a]According to the AJCC (American Joint Committee on Cancer) staging system [42].

doi:10.1371/journal.pone.0103270.t001

Purified cDNA library products were then evaluated using the Agilent bio analyzer and diluted to 10 nM for cluster generation in situ on the HiSeq paired-end flow cell using the CBot automated cluster generation system followed by massively-parallel sequencing (2×100 bp) on HiSeq 2000. We obtained 101 bp mate-paired reads from DNA fragments of an average length of 250 bp (standard deviation for the distribution of inner distances between mates pairs is approximately 50 bp). RNA-seq reads from cancer tissues and adjacent tissues were separately aligned to the human genome using the software TopHat (version.1.1.4).

Quantitative RT-PCR

The clinical characteristics of the breast cancer patients have been described in Table 1. Real-time quantitative RT-PCR was performed to determine gene expression in the samples. Total RNA was isolated using the Qiagen RNeasy kit. First strand cDNA was synthesized as the following: total of 1 μg of RNA from each sample was reverse-transcribed using random primers and M-MLV reverse transcriptase according to the protocol of the manufacturer. Quantitative PCR was performed using SYBR Premix Ex Taq. qPCR was done in triplicates in the ABI prism 7300 sequence detector. The relative amounts of gene expression were calculated (ΔΔCT method) by using the expression of β-actin as an internal standard. The formula based on the threshold cycle (Ct) [41] is as follows:

$$RQ = 2^{-\Delta\Delta Ct}$$

Here, ΔΔCt = (Ct lincRNA- Ct β-actin)cancer–(Ct lincRNA- Ct β-actin)adjacent.

General statistical analysis for qPCR

Real time qPCR was repeated at least in three independent experiments in every sample. Data were presented as mean±SD of three or more independent experiments. Statistical analysis was performed with SPSS (version17.0) and the differences were considered statistically significant when P value was less than 0.05 by using the independent samples t-test.

Supplementary Information

The result of the next sequencing was shown in Table S1, S2, S3, S4, and S9. The information of selected lincRNAs for validated was shown in the Table S5. The information of primers for RT-PCR and qPCR was shown in the Table S6 and Table S7. The result of correlations analysis was shown in the Table S8.

Supporting Information

Table S1 The 124 lincRNAs exclusively expressed in adjacent tissues to cancer by deep sequencing.
(XLS)

Table S2 The 62 lincRNAs exclusively expressed in the cancer tissues.
(XLS)

Table S3 The up-expressed lincRNAs in breast cancer tissues compared with adjacent tissues.
(XLS)

Table S4 The down-expressed lincRNAs in breast cancer tissues compared with adjacent tissues.
(XLS)

Table S5 The information of selected lincRNAs for semi-quantitative PCR.
(XLS)

Table S6 The primers information of lincRNAs for RT-PCR.
(XLS)

Table S7 The primers information of lincRNA for qPCR.
(XLS)

Table S8 The results of data from real-time PCR in breast tissues analyzed by RQ and SPSS 17.0.
(XLS)

Table S9 The total aberrantly expression lincRNAs in breast cancer tissues by deep sequencing.
(XLS)

Author Contributions

Conceived and designed the experiments: XD HY. Performed the experiments: TJ LZ. Analyzed the data: XZ FW SG HY. Contributed reagents/materials/analysis tools: HY. Wrote the paper: XD LZ TJ. Revised the manuscript: XD LZ. Gave final approval of the version to be published: XD. Sample collection: MZ.

References

1. Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799–816.
2. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, et al. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proceedings of the National Academy of Sciences 106: 11667–11672.
3. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, et al. (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464: 1071–1076.
4. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N (1996) Requirement for Xist in X chromosome inactivation. Nature 379: 131–137.
5. Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, et al. (2012) LincRNA-p21 suppresses target mRNA translation. Molecular cell.
6. Ji P, Diederichs S, Wang W, Böing S, Metzger R, et al. (2003) MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene 22: 8031–8041.
7. He JH, Han ZP, Li YG (2014) Association between long non-coding RNA and human rare diseases (Review). Biomedical Reports 2: 19–23.
8. Huarte M, Rinn JL (2010) Large non-coding RNAs: missing links in cancer? Human molecular genetics 19: R152–R161.
9. Siegel R, Naishadham D, Jemal A (2013) Cancer statistics, 2013. CA: a cancer journal for clinicians 63: 11–30.
10. Calin GA, Croce CM (2006) MicroRNA signatures in human cancers. Nature Reviews Cancer 6: 857–866.
11. Wapinski O, Chang HY (2011) Long noncoding RNAs and human disease. Trends in cell biology 21: 354–361.
12. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10: 57–63.
13. Metzker ML (2009) Sequencing technologies—the next generation. Nature Reviews Genetics 11: 31–46.
14. Hawkins RD, Hon GC, Ren B (2010) Next-generation genomics: an integrative approach. Nature Reviews Genetics 11: 476–486.
15. Washietl S (2010) Sequence and structure analysis of noncoding RNAs. Data Mining Techniques for the Life Sciences: Springer. pp. 285–306.
16. Bernhart SH (2011) RNA Structure Prediction. In Silico Tools for Gene Discovery: Springer. pp. 307–323.
17. Muppirala UK, Honavar VG, Dobbs D (2011) Predicting RNA-protein interactions using only sequence information. BMC bioinformatics 12: 489.
18. Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9: 387–402.
19. Suzuki T, Higgins P, Crawford D (2000) Control selection for RNA quantitation. Biotechniques 29: 332–337.
20. Dhillon VS, Husain SA, Ray G (2003) Expression of aphidicolin-induced fragile sites and their relationship between genetic susceptibility in breast cancer, ovarian cancer, and non-small-cell lung cancer patients. Teratogenesis, carcinogenesis, and mutagenesis 23: 35–45.
21. Richard F, Pacyna-Gengelbach M, Schlüns K, Fleige B, Winzer KJ, et al. (2000) Patterns of chromosomal imbalances in invasive breast cancer. International journal of cancer 89: 305–310.
22. Pastural E, Ersoy F, Yalman N, Wulffraat N, Grillo E, et al. (2000) Two genes are responsible for Griscelli syndrome at the same 15q21 locus. Genomics 63: 299–306.
23. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M, et al. (2010) Common variants at 2q37. 3, 8q24. 21, 15q21. 3 and 16q24. 1 influence chronic lymphocytic leukemia risk. Nature genetics 42: 132–136.
24. Liu YZ, Xiao P, Guo YF, Xiong DH, Zhao LJ, et al. (2006) Genetic linkage of human height is confirmed to 9q22 and Xq24. Human genetics 119: 295–304.
25. Öhman M, Oksanen L, Kaprio J, Koskenvuo M, Mustajoki P, et al. (2000) Genome-wide scan of obesity in Finnish sibpairs reveals linkage to chromosome Xq24. Journal of Clinical Endocrinology & Metabolism 85: 3183–3190.
26. Melchor L, Saucedo-Cuevas LP, Muñoz-Repeto I, Rodríguez-Pinilla SM, Honrado E, et al. (2009) Comprehensive characterization of the DNA amplification at 13q34 in human breast cancer reveals TFDP1 and CUL4A as likely candidate target genes. Breast Cancer Res 11: R86.
27. Shinomiya T, Mori T, Ariyama Y, Sakabe T, Fukuda Y, et al. (1999) Comparative genomic hybridization of squamous cell carcinoma of the esophagus: the possible involvement of the DP1 gene in the 13q34 amplicon. Genes, Chromosomes and Cancer 24: 337–344.
28. Dohna M, Reincke M, Mincheva A, Allolio B, Solinas-Toldo S, et al. (2000) Adrenocortical carcinoma is characterized by a high frequency of chromosomal gains and high-level amplifications. Genes, Chromosomes and Cancer 28: 145–152.
29. Michiels EM, Weiss MM, Hoovers JM, Baak JP, Voute P, et al. (2002) Genetic alterations in childhood medulloblastoma analyzed by comparative genomic hybridization. Journal of pediatric hematology/oncology 24: 205–210.
30. Yasui K, Arii S, Zhao C, Imoto I, Ueda M, et al. (2002) TFDP1, CUL4A, and CDC16 identified as targets for amplification at 13q34 in hepatocellular carcinomas. Hepatology 35: 1476–1484.
31. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, et al. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science 329: 689–693.
32. Cirillo D, Agostini F, Tartaglia GG (2012) Predictions of protein–RNA interactions. Wiley Interdisciplinary Reviews: Computational Molecular Science.
33. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266: 66–71.
34. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, et al. (1994) Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. Science 265: 2088–2090.
35. Osborne CK (1998) Steroid hormone receptors in breast cancer management. Breast cancer research and treatment 51: 227–238.
36. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, et al. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235: 177–182.
37. Bos JL (1989) Ras oncogenes in human cancer: a review. Cancer research 49: 4682–4689.
38. Li J, Yen C, Liaw D, Podsypanina K, Bose S, et al. (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 275: 1943–1947.
39. Balkwill F (2006) TNF-α in promotion and progression of cancer. Cancer and Metastasis Reviews 25: 409–416.
40. Osborne C, Wilson P, Tripathy D (2004) Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and therapeutic applications. The oncologist 9: 361–377.
41. Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nature protocols 3: 1101–1108.
42. Edge SB, Compton CC (2010) The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Annals of surgical oncology 17: 1471–1474.